

(19)日本国特許庁(J P)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-12351

(43)公開日 平成5年(1993)1月22日

(51)Int.Cl. ⁵	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 15/42		D 7060-5L		
15/18		8945-5L		
15/40	5 0 0 Q	7060-5L		
G 0 6 G 7/60		7368-5B		

審査請求 未請求 請求項の数1(全 17 頁)

(21)出願番号 特願平3-161819
(22)出願日 平成3年(1991)7月2日

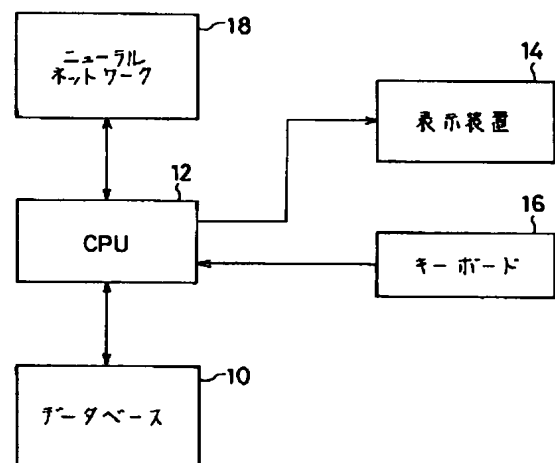
(71)出願人 000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者 大橋 昭南
東京都港区芝浦一丁目1番1号 株式会社
東芝本社事務所内
(74)代理人 弁理士 鈴江 武彦

(54)【発明の名称】 診断支援システム

(57)【要約】

【目的】本発明は、膨大な量の医療データを効率的に整理し、診断支援システムに組み込むことができる診断論理、知識を得ることを目的とする。

【構成】診断支援システムは、多数の医療データを記憶するデータベース10と、データベース10に記憶されている多数の医療データをLVQ学習方式により分類するニューラルネットワーク18と、入力されたサンプルデータを分類された医療データと比較して、データベース内の過去の医療データとの類似度を求め、類似データを検索したり、データ分類毎のデータの発生頻度を求め、それを提示したり、データ分類毎のデータの発生頻度に基づき、サンプルデータに応じた診断を導出するCPU12とを具備する。



【特許請求の範囲】

【請求項1】 多数の医療データを記憶する手段と、前記記憶手段に記憶されている多数の医療データをベクトル量子化学習により分類するニューラルネットワーク手段と、入力されたサンプルデータを分類された医療データと比較して、診断支援情報を出力する手段とを具備する診断支援システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、診断支援システムに関し、とくに診断論理の抽出に関する。

【0002】

【従来の技術】 診断支援システム(Computer-Aided-Diagnosis も含む)を構築する場合、その中に診断論理(あるいは知識)を組み込むことが必要であり、組み込み方法としては、(1)手続き型のプログラムで組み込む方法、(2)ルールベース、ファジィ論理等により組み込む方法、(3)ニューラルネットワークにより組み込む方法等が考えられる。一方、診断論理、知識を得るための方法としては以下の例が考えられている。

- (1) 専門家にインタビューして診断論理、知識を引き出す。
- (2) ニューラルネットワークにより学習する。
- (3) マシン・ラーニングにより知識を抽出する。
- (4) 多変量解析などの統計的手法により診断論理を抽出する。
- (5) 出現頻度による方式。

【0003】 しかし、これらの知識獲得方法はいずれも確立した方法ではない。第1の方法では、専門家は知識を経験的に得ている場合がほとんどであるので、それを体系的に整理し、計算機の理解する言語に変換することは非常に難しい。第2、第3の方法では、データと結果だけから計算機が知識を抽出するので、部分的には有効な方法であるが、知識が複雑になると、よい結果が得られるとは限らない。そのため、現在は研究段階であり、実用されていない。第4の方法でも、データと結果だけから知識を抽出するが、処理が線形であるので、非線形な知識の抽出が困難である。第5の方式としては、特開昭62-220868号公報に記載の自動分析装置があるが、この従来例は非常に簡単な方法であり、精度が不十分である。このように、従来の獲得方法では正しく診断論理を作成することが困難である。これは、膨大な量の医療データが効率的に整理されていないからである。

【0004】

【発明が解決しようとする課題】 本発明は上述した事情に対処すべくなされたもので、その目的は医療データを整理することができる診断支援システムを提供することである。

【0005】

【課題を解決するための手段】 本発明による診断支援システムは、多数の医療データを記憶するデータベースと、データベースに記憶されている多数の医療データを学習ベクトル量子化により分類するニューラルネットワークと、入力されたサンプルデータを分類された医療データと比較して、診断支援情報を出力する手段とを具備する。

【0006】

【作用】 本発明による診断支援システムによれば、医療データのある範囲に分類することができるので、このデータ分類に応じてデータベース内の過去の医療データとの類似度を求め類似データを検索することや、データ分類毎のデータの発生頻度を求め、それを提示することや、データ分類毎のデータの発生頻度に基づき、サンプルデータに応じた診断を導出することができる。

【0007】

【実施例】 以下、図面を参照して本発明による診断支援システムの実施例を説明する。ここでは、肝臓の疾患に関する判定を行なう健康診断システムに適用した実施例を説明する。

【0008】 先ず、このような健康診断システムの概要を説明する。説明の簡単化のために、診断に使う1つの医療データのデータ項目はGOT値、GPT値、 γ -GTP値の3種類とし、頻度を求めるための過去のデータ数を約3000例とするが、データの種類の数はこれに限定されない。また、データ項目が3種類の場合でも、GOT値、GPT値、 γ -GTP値以外の3種類であってもよい。さらに、ここでは、データは数値として得られるとしたが、例えば画像データ等の場合に、その特徴を数値に変換したデータであってもよい。画像データの特徴を数値に変換する方法としては、(1)専門家が画像を観察してその特徴を数値に変換する方法、(2)専門家が画像上で指定したROI(関心領域)等の所定の部分の特徴をアルゴリズムで求めて数値に変換する方法、(3)計算機がパターン認識を行い画像の特徴を数値に変換する方法等があるが、これらに限定されない。

【0009】 次に、健康診断システムにおける判定について説明する。健康診断システムでは最終的な診断も重要であるが、患者の生活指導のために医師が以下の判定を行なっている。

- A：正常である。
- B：僅かに異常が認められるが、日常生活に差し支えない。
- BF：Bであるが、経過観察を要する。
- C：日常生活上、注意を要する。
- D：治療を要する。
- G：精密検査、または再検査を要する。

さらに、説明を簡単にするために、判定C、D、Gをまとめて扱い、これを判定CDGと称する。この実施例で取り扱う検査項目は、全ての検査が行われており、全て

の検査データが揃っており、欠落データはないものとする。

【0010】以下の説明で使われる記号を定義する。h は判定結果を示す添え字であり、h = 1 は判定A、h = 2 は判定B、h = 3 は判定BF、h = 4 は判定CDGを示す。j はデータ項目を示す添え字であり、j = 1 はGOT値、j = 2 はGPT値、j = 3 はγ-GTP値を示す。

【0011】本実施例はベクトル量子化学習 (Learning Vector Quantization: 以下LVQと称する) 方式により学習を行なうニューラルネットワークを用いて医療データを整理しており、次にこのLVQ方式の学習について簡単に説明する。ニューラルネットワークの学習方式にはホップフィールド (Hopfield) モデル、バックプロパゲーション方式などがよく知られているが、その1つにLVQ方式がある。ここでは、説明を簡単にするために、ニューラルネットワークが対象とするデータは2次元データであるとする。

【0012】LVQ方式とは図2に示すように、あるデータの分布を別の分布として写像する方式である。LVQ方式の学習を行なうニューラルネットワークは、図3に示すように2次元配列されている素子を有する。各素子毎に位置 (p, q) が決まっている。この素子数は、対象とする問題に応じて適宜選択される。ここでは、4 × 5 個の素子が選択されている。各素子をNUp,q (p = 1 ~ 4, q = 1 ~ 5) で表す。各素子NUp,q はそれぞれ対象とするデータの次元数だけのベクトルを持っている。ここでは対象とするデータは2次元であるので、素子NUp,qは2次元のベクトルを持っている。そのため、素子のベクトルをNUp,q (Xp, Yq) で表す。LVQ方式の学習の手順を図4に示すフローチャートを参照して説明する。

【0013】ステップ#1で、各素子NUp,q (Xp, Yq) に適当な初期値、例えば乱数を設定する。ただし、 $0 \leq Xp \leq 1$, $0 \leq Yq \leq 1$ である。ステップ#2で、学習半径LRに初期値を設定する。ステップ#3で、1サイクルの学習毎の学習半径LRの減少量を指定する。ステップ#4で、学習サイクル数を指定する。

【0014】ステップ#5～ステップ#8で、1つの学習データについての学習を行なう。まず、ステップ#5で、1つの学習データと、全ての素子データNUp,q (Xp, Yq) とのユークリッド距離を求める。ここでは20個の素子があるので、20個の距離を求める。ステップ#6で、この距離のなかで最も短い距離を持つ素子を探し、これを勝者素子NUcp,cqとする。ステップ

$$NDVj = 1 / [1 + \exp \{-NKAj (RDVj - MTVj)\}] \quad (1)$$

ここで、 $NKAj = -1 / \ln 0.25 / (DMXj - MTVj)$ 、 \ln は自然対数である。

【0019】(1)式により正規化を行えば、データが最大値DMXjの場合に0.8に、標準的な値MTVj

#7で、勝者素子のベクトル (Xcp, Ycq) をこの学習データのベクトルに近付く (一致する) ように変更する。変更量はデータの各ベクトル成分 (各項目) 毎の差に比例した量である。ステップ#8で、勝者素子の位置 (Cp, Cq) からの距離が、学習半径LR内にある全ての素子についても、勝者素子と同様に各素子のベクトルを学習データのベクトルに近付くように変更する。

【0015】ステップ#9で、全ての学習データについて学習が行なわれたか否かを判定する。全ての学習データについての学習が終わっていない場合は、ステップ#5に戻る。全ての学習データについての学習が終わった場合は、1サイクルの学習が終了したと判断し、ステップ#10で、学習半径LRを指定された減少量だけ減少させる。

【0016】ステップ#11で、学習サイクル数がステップ#4で指定された回数に達したか否かを判定する。指定された回数だけ学習サイクルが実行されていない場合は、ステップ#5に戻る。指定された回数の学習サイクルが実行された場合は、LVQ学習を終了する。

【0017】このようなLVQ学習により、学習データの分布がニューラルネットワーク上に分類されて写像される。そのため、学習後のニューラルネットワークに、あるサンプルデータを入力すると、ニューラルネットワークはそのデータに応じて各素子からデータを出力する。出力データはそのデータのベクトルと各素子のベクトルとの距離を入出力関数により変換した値であり、入出力関数としては、例えば $1 / \{1 + (\text{距離の2乗})\}$ のような関数が使われる。従って、各素子の出力データは距離が0、すなわち一致する場合に1 (最大値) になり、距離が遠くなるにつれて0 (最小値) に近づく。このように、ニューラルネットワークは、入力データの写像を、各素子の出力データ (0 ~ 1) の分布として表わすことができる。

【0018】次に、データの正規化について説明する。データはベクトルデータであり、複数の成分を有し、各成分の取り得る値の範囲はそれぞれに異なるので、データを項目毎に正規化することが好ましい。正規化は例えば以下のように行う。データ項目j毎に過去のデータから標準的な値を求め、これをMTVjとする。標準的な値は例えば、平均値あるいは中央値あるいは最頻度値などである。また、最大値を求め、これをDMXjとする。あるデータの測定値をRDVj、正規化後のデータの値をNDVjとすると、例えば、次のようなシグモイド関数により、測定値は正規化される。

の場合には0.5に正規化され、全てのデータは0 ~ 1に正規化される。以下、データは全て0 ~ 1に正規化されているものとして説明する。

【0020】次に、第1実施例の診断支援システムを説

明する。図1は第1実施例の構成を示すブロック図である。本実施例は、過去の多数の医療データ（ここでは、1データは3次元のベクトルデータであり、GOT値、GPT値、 γ -GTP値の3つの成分からなる）が格納されているデータベース10、データベース10からデータを検索するための中央処理ユニット（CPU）12、データベース10のデータをLVQ学習方式により分類するためのニューラルネットワーク18、検索データや検索を行なう際のガイダンス等を表示する表示装置14、CPU12に種々の指示を与えたり、データを入力するためのキーボード16からなる。

【0021】本実施例は与えられたサンプルデータに類似する過去のデータをデータベース10より検索する方式に関するものであり、ニューラルネットワーク18のLVQ学習方式による写像を利用することにより、データベース10内の医療データのある範囲毎に分類し、分類後の写像空間において、サンプルデータとデータベース10内のデータとの類似性を求め、その類似性に応じ

$$\sum (NNV_{xj} - NDV_{xj})^2$$

【0023】しかし、ユークリッド距離はデータの特性をなんら考慮していないので、ばらつきの大きいデータの影響が大きくなる。このために、一般的にはデータの

$$\sum WGH_j (NNDV_{xj} - NDV_{xj})^2$$

【0024】しかしながら、実際にはこの重み WGH_j を決めることが困難であるので、正しく類似度を求めることが困難である。重み WGH_j を決める1つの方法は統計の手法を用いることである。そのために、種々の統計手法が用いられているが、確立した方法はない。これは、統計は線形処理がほとんどあるためであると考えられる。

【0025】そこで、本実施例はLVQ学習方式のニューラルネットワークの写像を利用することにより、医療データのある範囲毎に分類（写像）し、写像後の空間においてデータ間の類似度を定義する。本実施例の動作を図5に示したフローチャートを参照して説明する。ステップ#21で、ニューラルネットワーク18の素子構成を決める。ここでは、例えば、図3に示したように、4×5個の素子構成とする。

【0026】ステップ#22で、データの項目数を決める。ここでは、例えば、GOT値、GPT値、 γ -GTP値の3項目とする。ステップ#23で、データベース10内の全データについて、図4に示したようなニューラルネットワーク18を用いたLVQ方式の学習を実行する。

ここで、

$$SVE_n^2 = \sum \sum WGH_{p,q} \times (NUTS_{p,q} - NUTn_{p,q})^2 \quad (5)$$

$WGH_{p,q}$ は素子の出力に対する重みであり、 $WGH_{p,q} = WGHF1(NUTS_{p,q}) + WGHF2(NUT$

データベース10より類似データを検索することを目的としている。

【0022】図1において、データベース10内に過去の3000例の医療データがあるとする。操作者がある患者についての上記3項目のデータGOT値、GPT値、 γ -GTP値からなるサンプルデータを入力し、それに類似した過去の医療データを取り出す動作を説明する。この場合、入力したサンプルデータとまったく一致したデータがデータベース10内に存在する確率は非常に低い。従って、過去のデータの中から、最も類似したデータを探すことが要求される。そのために、データの類似度を定義する必要がある。類似度の最も簡単な定義は、データ間のユークリッド距離（正規化した）である。データベース10内の過去の1つの医療データを NDV_x （ $NDV_{x1}, NDV_{x2}, \dots, NDV_{xj}, \dots$ ）、サンプルデータを $NNDV_x$ （ $NNDV_{x1}, NNDV_{x2}, \dots, NNDV_{xj}, \dots$ ）とすると、2個のデータ間のユークリッド距離の2乗は下記により求まる。

$$(2)$$

項目毎に重み付けをするのが一般的である。項目毎の重みを WGH_j とすると、重み付きユークリッド距離の2乗は下記により求まる。

$$(3)$$

【0027】ステップ#24で、過去の医療データの中から類似データを検索するための新しいサンプルデータをニューラルネットワーク18に入力し、各素子の出力を求める。サンプルデータに対する素子出力を $NUTS_{p,q}$ とする。ここで、 p, q は素子の番号であり、 $p = 1 \sim 4, q = 1 \sim 5$ である。

【0028】ステップ#25～ステップ#27で、データベース10の全データとサンプルデータとの類似度を求める。データベース10のデータを NDV_n （ $n=1, \dots$ ）とする。まず、ステップ#25で、データベース10中の1個のデータをニューラルネットワーク18に入力し、各素子の出力を求める。データベースのデータに対する素子出力を $NUTn_{p,q}$ とする。ステップ#26で、両データ間の類似度を求める。ここでは、次のように表わされる類似度 SV_n とする。ただし、類似度はLVQの写像後の空間におけるものであれば、この式に限定されることはない。

$$(4)$$

n, p, q) である。ここで、WGHF1、WGHF2はNUTSp,q、NUTn,p,qに関する重みを決めるための関数である。例えば、サンプルデータに対する素子出力を重視し、かつ出力値が高いものを重視して、

$$\begin{aligned} \text{WGHF1}(\text{NUTSp,q}) &= \text{NUTSp,q} \\ \text{WGHF2}(\text{NUTn,p,q}) &= 0 \end{aligned}$$

とすれば、

$$\begin{aligned} \text{SVEn} &= \sum \sum \text{NUTSp,q} \times (\text{NUTSp,q} - \text{NUTn,p,q})^2 \quad (6) \end{aligned}$$

【0029】となる。また、 $\sum_{p=1}^{p-1} \sum_{q=1}^{q-1}$ サンプルデータに対する素子出力NUTSp,qが最高値である素子の番号をpq1、

以下大きい順にpqm(m=1～20)とすると、例えば、上位5番目までを重視して、

$$\text{SVEn} = \sum_{m=1}^5 \text{NUTSpqm} \times (\text{NUTSpqm} - \text{NUTn,pqm})^2 \quad (7)$$

とすれば、計算が少なく済む。さらに、最高値の素子

だけに注目すれば、

$$\text{SVEn} = \text{NUTSpq1} \times (\text{NUTSpq1} - \text{NUTn,pq1})^2 \quad (8)$$

となり、さらに簡単になる。

【0030】ステップ#27で、データベース10中の全てのデータとサンプルデータとの類似度が求められたか否かを判定する。全てのデータについての類似度が求められていない場合は、ステップ#25に戻る。全てのデータについての類似度が求められた場合は、ステップ#28で、所定個数のデータ、例えば、最大から5番目までの上位5つの類似度を持つデータベース10中のデータをサンプルデータに対する類似データとして提示する。その際、類似度SVも合わせて提示するようにするとさらによい。

【0031】以上説明したように本実施例によれば、ニューラルネットワークのLVQ学習によりデータベース10内の多数のデータの特性が分類できるので、データの特性に応じた類似度を持つデータを正しく検索することが可能であり、有効な診断支援情報が得られる。

【0032】次に、第2実施例について説明する。本実施例はニューラルネットワークのLVQ学習を利用して、医療データのある範囲に分類し、その分類毎にデータの発生頻度を求め、サンプルデータに対応した発生頻度を提示するとともに、発生頻度に応じた診断を導出

し、提示するデータベースを提供することを目的としている。本実施例の構成は図1に示した第1実施例と同様であるので、説明は省略する。本実施例の動作を図6に示したフローチャートを参照して説明する。

【0033】ステップ#31で、ニューラルネットワーク18の素子構成を決める。ここでは、例えば、図3に示したように、4×5個の素子構成とする。ステップ#32で、データの項目数を決める。ここでは、例えば、GOT値、GPT値、γ-GTP値の3項目とする。ステップ#33で、データベース10内の全データについて、図4に示したようなニューラルネットワーク18を用いたLVQ方式の学習を実行する。

【0034】ステップ#34で、CPU12内のメモリに表1に示すような頻度表を作成し、記憶する。ここで、p、qはLVQによる分類の番号であり、H(1,p,q)、H(2,p,q)、H(3,p,q)、H(4,p,q)はそれぞれ分類(p、q)における判定A、判定B、判定BF、判定CDGの頻度を示す。なお、初期値としては全ての頻度に0が設定されている。

【0035】

【表1】

分類		H (1, p, q)	H (2, p, q)	H (3, p, q)	H (4, p, q)
p	q	判定 A	判定 B	判定 B F	判定 C D G
1	1				
	2				
	3				
	4				
	5				
2	1				
	2				
	3				
	4				
	5				
3	1				
	2				
	3				
	4				
	5				
4	1				
	2				
	3				
	4				
	5				

データベース10のデータの番号をn、全データ数をDNとし、データベース10のデータをNDVn (n=1~DN) とする。まず、n=1である。

【0036】ステップ#35で、データベース10中のデータNDVn をニューラルネットワーク18に入力し、各素子の出力を求める。データベースのデータに対する素子出力をNUTn, p, q とする。ステップ#36で、出力NUTn, p, q が最大である素子を探し、その素子の番号をpm, qm とする。この時、このデータNDVn は分類 (pm, qm) に属すると定義する。データベース10内にはデータNDVn に関して医師が行なった判定もデータとともに記憶されており、ステップ#37でこの判定 (j=1~4) を読出し、これをDh とする。ステップ#38で、頻度H (Dh, pm, qm) をインクリメント (+1) して (H (Dh, pm, qm) ← H (Dh, pm, qm) + 1)、頻度表を更新する。

【0037】ステップ#39でデータベース10内の全データについて処理を行なったか否かを判定する。すなわち、n=DNか否かを判定し、n=DNでなければ、nを1だけインクリメント (+1) して、ステップ#35に戻る。n=DNであれば、全データの処理が終了

$$\begin{aligned}
 & HH(spm, sqm) \\
 &= H(1, spm, sqm) + H(2, spm, sqm) + H(3, spm, sqm) + H(4, spm, sqm) \\
 & H\%(1, spm, sqm) = 100 \times H(1, spm, sqm) / HH(spm, sqm) \\
 & H\%(2, spm, sqm) = 100 \times H(2, spm, sqm) / HH(spm, sqm) \\
 & H\%(3, spm, sqm) = 100 \times H(3, spm, sqm) / HH(spm, sqm) \\
 & H\%(4, spm, sqm) = 100 \times H(4, spm, sqm) / HH(spm, sqm)
 \end{aligned}$$

【0041】なお、これらの結果が、医師の判断と大きく異なる場合は、LVQによるデータの分類が不十分で

し、データベース10内のデータについての頻度表が完成したと判断でき、次のステップに進む。

【0038】ステップ#40で、サンプルデータをニューラルネットワーク18に入力し、各素子の出力を求める。サンプルデータに対する素子出力をNUTSp, q とする。

【0039】ステップ#41で、出力NUTSp, q が最大である素子を探し、その素子の番号をspm, sqm とする。ここで、このサンプルデータは分類 (spm, sqm) に属すると定義する。

【0040】ステップ#42で、この分類に属するデータの頻度H (1, spm, sqm)、H (2, spm, sqm)、H (3, spm, sqm)、H (4, spm, sqm) をCPU12内の頻度表から読出し、これを提示する。なお、頻度H (1, spm, sqm)、H (2, spm, sqm)、H (3, spm, sqm)、H (4, spm, sqm) の中で最大値を持つ判定の名称A、B、BF、またはCDGも、併せて提示してもよい。さらに、以下のような頻度の合計HH (spm, sqm) に対する各判定毎の頻度の割合H% (1, spm, sqm)、H% (2, spm, sqm)、H% (3, spm, sqm)、H% (4, spm, sqm) をサンプルデータに対する各頻度の確信度として、併せて提示してもよい。

あるので、ニューラルネットワーク18の素子数を変更してLVQ学習をやり直すようにしてもよい。

【0042】以上説明したように本実施例によれば、ニューラルネットワークのLVQ学習によりデータベース10内の多数のデータの特性を分類した後、サンプルデータに対して類似したデータに対する医師の過去の確定診断を頻度として提示することができ、医師の判断の手助けとなる診断支援情報が得られる。

【0043】次に、第3実施例について説明する。本実施例の原理は、LVQ学習を利用して第1のニューラルネットワークで医療データのある範囲に分類（写像）し、その分類と医師の診断の関係を第2のニューラルネットワークで学習して、診断結果を導出し、提示することを目的としている。

【0044】全体の構成を示すブロック図は図1と同じであるので、図示省略するが、ニューラルネットワークの構成は図7に示すような3層の階層型である。入力層18aは3素子であり、各素子は入力データの各項目、例えばGOT値、GPT値、 γ -GTP値にそれぞれ割り当てられている。各入力データは上述の実施例と同様に正規化されている。中間層18bは前述の実施例のニューラルネットワーク18と同様に、LVQ学習のための素子であり、ここでは前述の実施例と同様に、 4×5 の素子構成である。そのため、入力層18aと中間層18bとが前述の第1のニューラルネットワークに相当する。

【0045】出力層18cは、ここでは4素子であり、それぞれ、判定A、B、BF、CGDに対応している。出力層18cの素子の入出力関数はシグモイド関数である。中間層18bの20（ $=4 \times 5$ ）個の素子の出力は、出力層18cの4個の素子の入力に、それぞれ重みを介して接続され、1つのニューラルネットワークを構成している。従って、重みは80（ $=20 \times 4$ ）個存在する。重みの初期値は乱数等の値を設定する。中間層18bと出力層18cとからなるニューラルネットワークが前述の第2のニューラルネットワークに相当する。入力層18aの素子数、入力データの項目数、中間層18bの素子数、出力層の素子数18cは、問題ごとに異なり、ここに示したものに限定されないことはもちろんである。本実施例の動作を図8に示したフローチャートを参照して説明する。

【0046】ステップ#41で、ニューラルネットワーク18の素子構成を決める。ここでは、例えば、図7に示したように、3層の階層型で、中間層が20素子、出力層が4素子の構成とする。ステップ#42で、データの項目数（入力層の素子数）を決める。ここでは、例えば、GOT値、GPT値、 γ -GTP値の3項目とする。ステップ#43で、データベース10内の全データについて、図4に示したようなLVQ方式の学習を実行する。この学習に使われるのは、ニューラルネットワ

ク18の中の入力層18aと中間層18bである。

【0047】ステップ#44で、データベース10内のデータの1個をニューラルネットワーク18に入力して、出力層18cの出力を求める。ステップ#45で、教師データとしてこのデータに対する医師の判定A、B、BF、またはCGDを与える。ここで、医師の判定に対応する出力層の素子の教師データは1、他の素子のそれは0である。ステップ#46で、ニューラルネットワーク18の出力と教師データとの誤差に応じて、Hebbの学習則により、中間層18bと出力層18cとの間の重みを変更する。

【0048】ステップ#47でデータベース10内の全データについて処理を行なったか否かを判定する。全部のデータに関する処理が終了していなければステップ#44に戻る。全データの処理が終了した場合は、ステップ#48で、誤差が予め設定した所定値以下であるか否かを判定する。所定値以下でなければ、ステップ#44に戻り、学習を繰り返す。所定値以下の場合は、ステップ#49でサンプルデータをニューラルネットワーク18に入力し、出力層18cの各素子の出力を得る。そして、最大の出力値を出力している素子に対応する判定の名称A、B、BF、またはCGDを提示する。その際、各素子の出力値を各判定毎の確信度として、併せて提示するようにしてもよい。

【0049】なお、これらの結果が、医師の判断と大きく異なる場合は、LVQ学習によるデータの分類が不十分である場合であるので、ニューラルネットワーク18の素子数、具体的には中間層18bの素子数を変更して、LVQによる学習をやり直すようにしてもよい。

【0050】以上説明したように本実施例によれば、ニューラルネットワークのLVQ学習によりデータベース10内の多数のデータの特性が分類した後、サンプルデータに対する診断結果をニューラルネットワークから出力することができるので、有用な診断支援情報が得られる。

【0051】次に、第4実施例について説明する。本実施例はニューラルネットワークのLVQ学習を利用して、医療データのある範囲に分類し、その分類毎にデータベースを構築し、サンプルデータに対応した分類毎のデータベースより類似データを検索することを目的としている。

【0052】データベースよりサンプルデータに類似したデータを検索するためには、データの類似度を定義し、その類似度に基づき全データベースの中から類似データを求めればよい。しかしながら、データベースのデータ数が多大になり、さらに1個のデータの項目数が多くなると、検索に非常に時間がかかるという問題がある。そのために、予めデータベースを分類しておけば、検索時間が短くて済む。しかしながら、データベースのデータを分類することは困難である。

【0053】そこで、本実施例は、LVQ方式のニューラルネットワークの写像を利用することにより、医療データをある範囲に分類し、その分類毎にデータベースを構築する。

【0054】本実施例の構成は図1に示した第1実施例と同様であるので、説明は省略する。ただし、図1ではデータベースは1つとして示されているが、本実施例では分類毎のデータベースとされている。

【0055】図1において、データベース10内に過去の3000例の医療データが記憶されているとする。操作者がある患者についての上記3項目のデータGOT値、GPT値、 γ -GTP値からなるサンプルデータを入力し、それに類似した過去の医療データを検索する動

作を説明する。この場合、入力したサンプルデータと全く一致したデータがデータベース10内に存在する確率は非常に低い。従って、過去のデータの中から、最も類似したデータを検索することが要求される。そのため、データの類似度を定義する必要がある。

【0056】ここでは類似度をデータ間のユークリッド距離（正規化した）を用いて定義する。データベース10内の過去の1つの医療データをNDV_x（NDV_{x1}, NDV_{x2}, ..., NDV_{xj}, ...）、サンプルデータをNN DV_x（NNDV_{x1}, NNDV_{x2}, ..., NNDV_{xj}, ...）とすると、2個のデータ間のユークリッド距離LUの2乗は下記により求まる。

$$LU^2 = \sum (NNDV_{xj} - NDV_{xj})^2 \quad (9)$$

次に、類似度SUを下記により定義する。

$$SU = 1 / (1 + LU^2) \quad (10)$$

なお、類似度の定義はこれに限定されることはない。本実施例の動作を図9、図10に示したフローチャートを参照して説明する。

【0057】ステップ#51で、ニューラルネットワーク18の素子構成を決める。ここでは、例えば、図3に示したように、4×5個の素子構成とする。ステップ#52で、データの項目数を決める。ここでは、例えば、GOT値、GPT値、 γ -GTP値の3項目とする。ステップ#53で、データベース10内の全データについて、図4に示したようなニューラルネットワーク18を用いたLVQ方式の学習を実行する。データベース10のデータの番号をn、全データ数をDNとし、データベース10のデータをNDV_n（n=1~DN）とする。まず、n=1である。

【0058】ステップ#54で、データベース10中のデータNDV_nをニューラルネットワーク18に入力し、各素子の出力を求める。データベースのデータに対する素子出力をNUT_{n, p, q}とする。ステップ#55で、出力NUT_{n, p, q}が最大である素子を探し、その素子の番号をpm, qmとする。この時、このデータNDV_nは分類（pm, qm）に属すると定義する。

【0059】ステップ#56で、図11に示すように、データベース10内の分類（pm, qm）のデータベースにデータNDV_n（正規化前のデータ、医師の判定等も含む）を記憶する。

【0060】ステップ#57でデータベース10内の全データについて処理を行なったか否かを判定する。すなわち、n=DNか否かを判定し、n=DNでなければ、nを1だけインクリメント（+1）して、ステップ#54に戻る。n=DNであれば、全データの処理が終了し、データベース10内のデータについての分類毎のデータベースが完成したと判断でき、次のステップに進む。

【0061】ステップ#58で、サンプルデータをニューラルネットワーク18に入力し、各素子の出力を求める。サンプルデータに対する素子出力をNUT_{Sp, q}とする。

【0062】ステップ#59で、出力NUT_{Sp, q}が最大である素子を探し、その素子の番号をspm, sqmとする。ここで、このサンプルデータは分類（spm, sqm）に属すると定義する。

【0063】分類（pm, qm）のデータベースのデータの番号をi、全データ数をDIとし、分類（pm, qm）のデータベースのデータをNDV_{pm, qm, i}（i=1~DN）とする。まず、i=1である。

【0064】ステップ#60～ステップ#63で、分類（pm, qm）のデータベースの全データとサンプルデータとの類似度を求める。まず、ステップ#60で、分類（pm, qm）のデータベースより1個のデータNDV_{pm, qm, i}を読み出す。ステップ#61で、サンプルデータNNDV_xとデータNDV_{pm, qm, i}との間のユークリッド距離LU_iの2乗を第9式により求める。ステップ#62で、両データ間の類似度SU_iを求める。

【0065】ステップ#63で、分類（pm, qm）のデータベース中の全てのデータとサンプルデータとの類似度が求められたか否かを判定する。全てのデータについての類似度が求められていない場合は、ステップ#60に戻る。全てのデータについての類似度が求められた場合は、ステップ#64で、所定個数のデータ、例えば、最大から5番目までの上位5つの類似度を持つ分類（pm, qm）のデータベース中のデータをサンプルデータに対する類似データとして提示する。その際、類似度SUも合わせて提示するようにするとさらによい。

【0066】以上説明したように本実施例によれば、ニューラルネットワークのLVQ学習によりデータベース10内の多数のデータを分類して、分類毎のデータベ

スを構築できるので、類似データを短時間に検索することが可能になる。

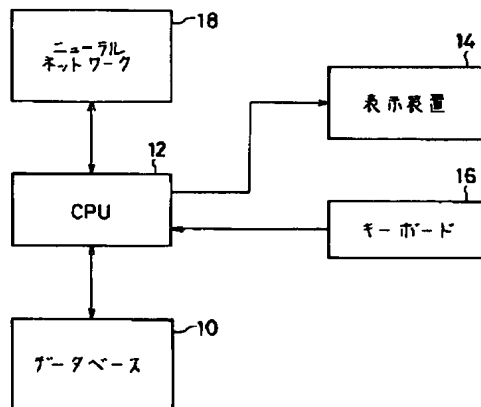
【0067】なお、本実施例により分類毎のデータベースを構築した後に、第2実施例の発生頻度を求めるようにしてもよい。さらに、本実施例による類似データと発生頻度の両方を提示するようにしてもよい。

【0068】なお、本発明は上述した実施例に限定されず、種々変形して実施可能である。例えば、肝臓の疾患に関する判定を行なう健康診断システムを説明したが、これに限定されない。

【0069】

【発明の効果】以上説明したように本発明によれば、医療データのある範囲に分類することができるので、このデータ分類に応じてデータベース内の過去の医療データとの類似度を求め、類似データを検索することや、データ分類毎のデータの発生頻度を求め、それを提示することや、データ分類毎のデータの発生頻度に基づき、サンプルデータに応じた診断を導出することができる診断支援システムが提供される。さらに、多数の医療データの各データとそれに対する医師の診断結果から、診断論理が得られるので、診断コンサルテーションシステムも容易に実現できる。

【図1】



【図面の簡単な説明】

【図1】本発明による診断支援システムの第1実施例の構成を示すブロック図。

【図2】LVQ学習による写像の原理を示す図。

【図3】LVQ学習を行なうニューラルネットワークの素子構成を示す図。

【図4】LVQ学習の手順を示すフローチャート。

【図5】第1実施例の動作を示すフローチャート。

【図6】本発明による診断支援システムの第2実施例の動作を示すフローチャート。

【図7】本発明による診断支援システムの第3実施例のニューラルネットワークの構成を示す図。

【図8】第3実施例の動作を示すフローチャート。

【図9】第4実施例の動作の前半を示すフローチャート。

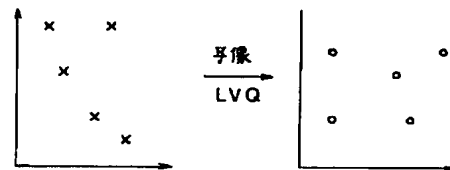
【図10】第4実施例の動作の後半を示すフローチャート。

【図11】第4実施例によるデータベースの分類を説明するための概略図。

【符号の説明】

10…データベース、12…CPU、14…表示装置、16…キーボード、18…ニューラルネットワーク。

【図2】

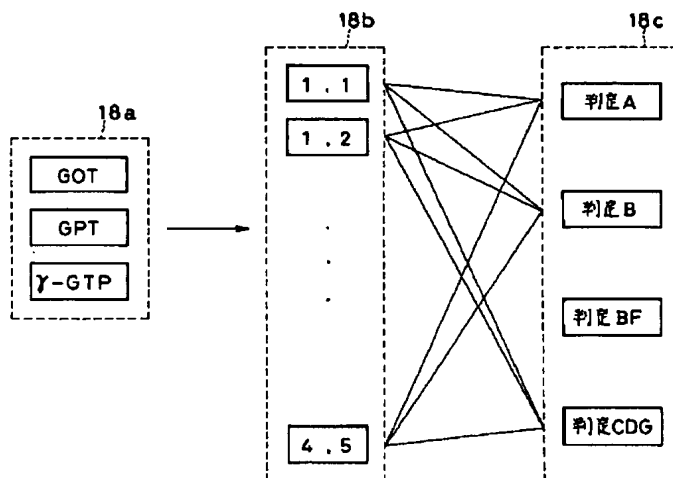


【図3】

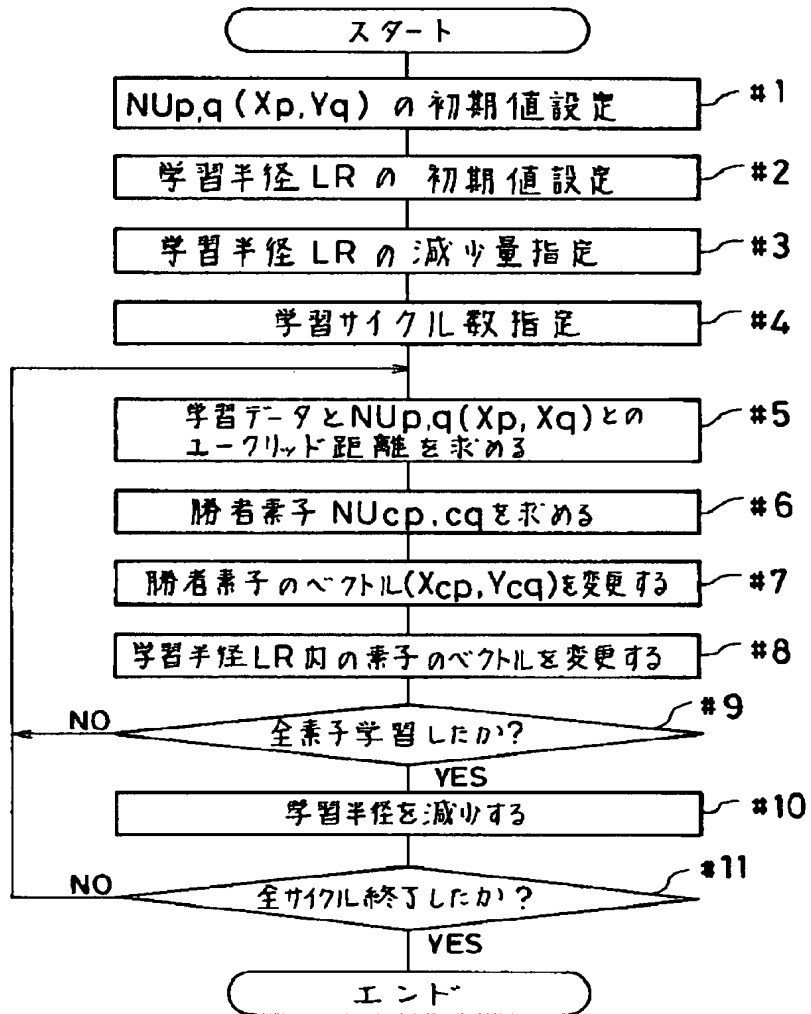
5	(X_1, Y_5)	(X_2, Y_5)	(X_3, Y_5)	(X_4, Y_5)
4	(X_1, Y_4)	(X_2, Y_4)	(X_3, Y_4)	(X_4, Y_4)
3	(X_1, Y_3)	(X_2, Y_3)	(X_3, Y_3)	(X_4, Y_3)
2	(X_1, Y_2)	(X_2, Y_2)	(X_3, Y_2)	(X_4, Y_2)
1	(X_1, Y_1)	(X_2, Y_1)	(X_3, Y_1)	(X_4, Y_1)
	1	2	3	4

p

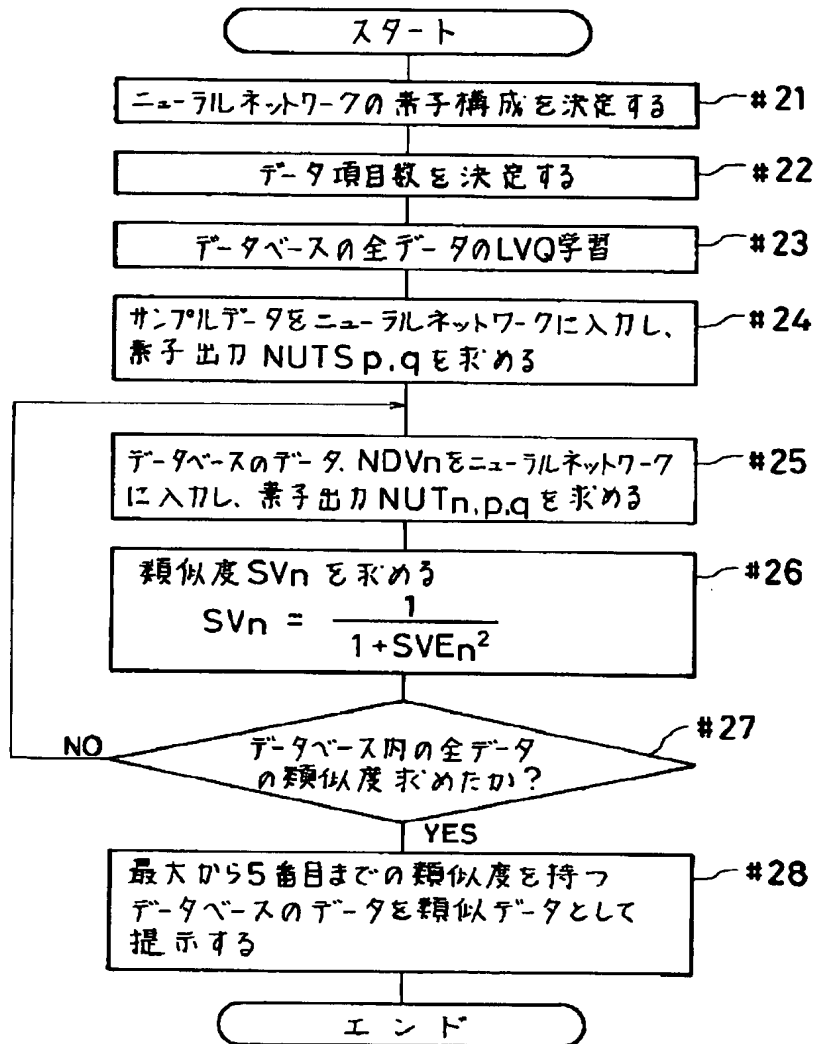
【図7】



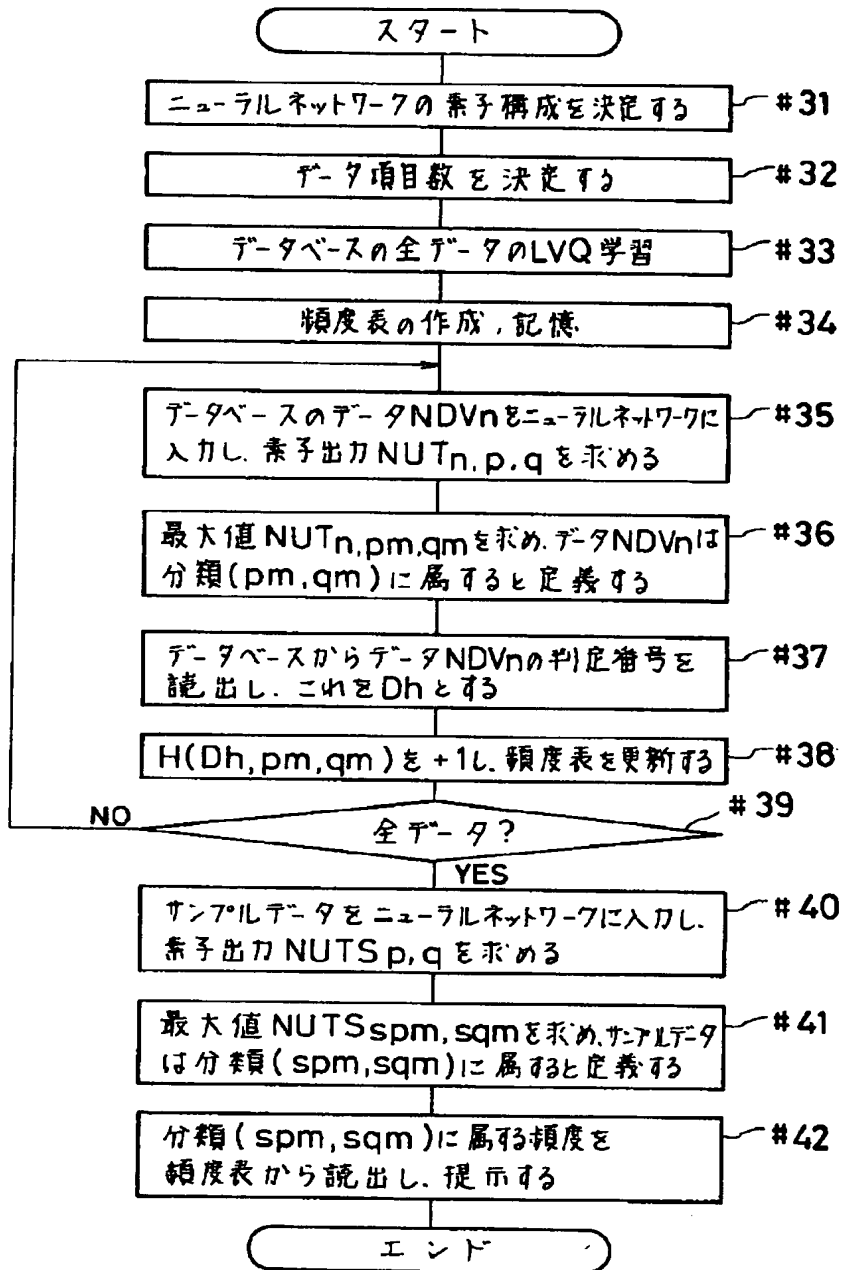
【図4】



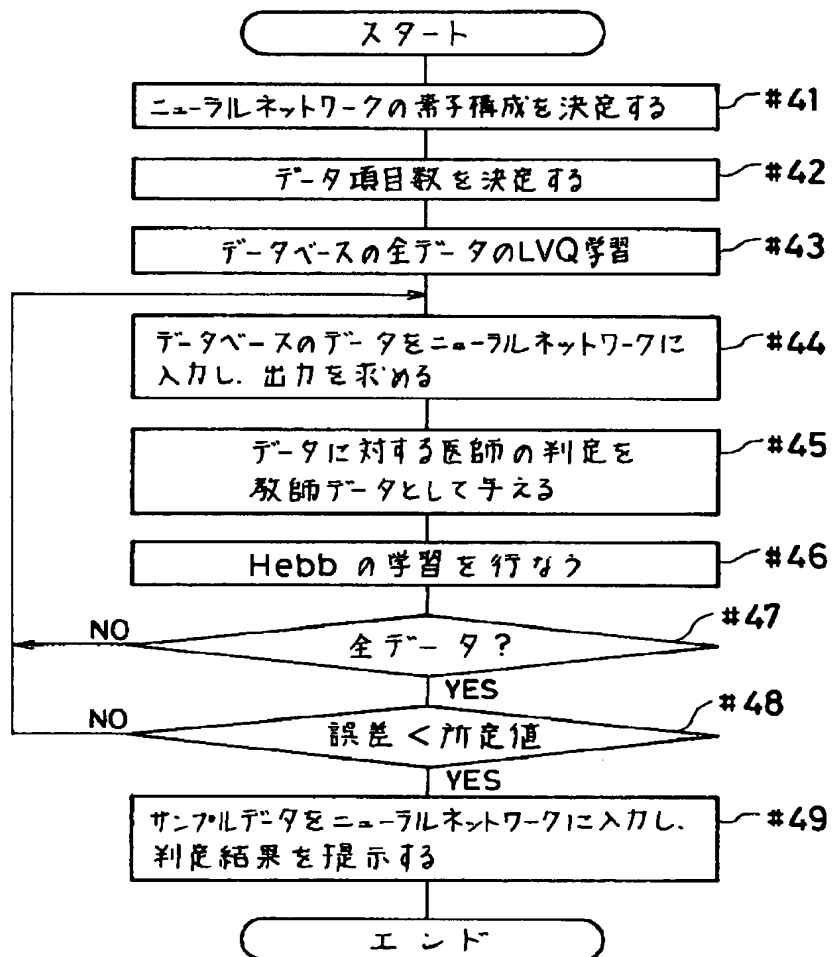
【図5】



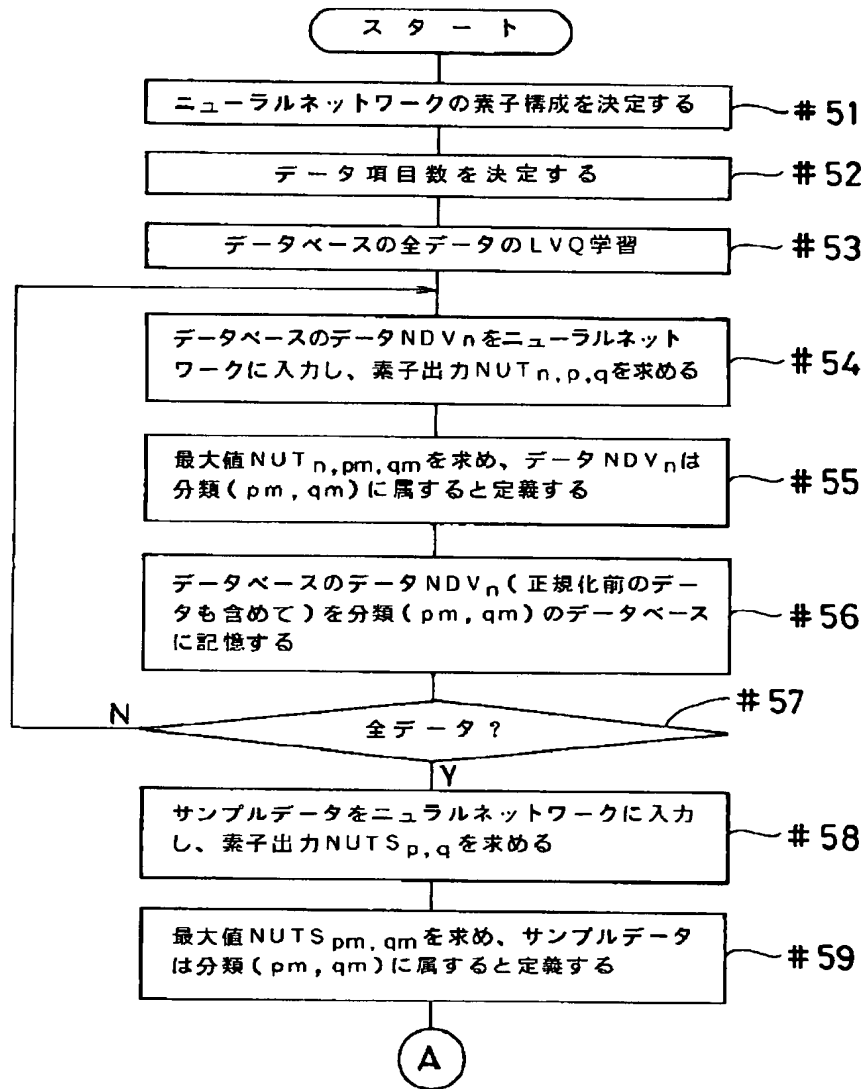
【図6】



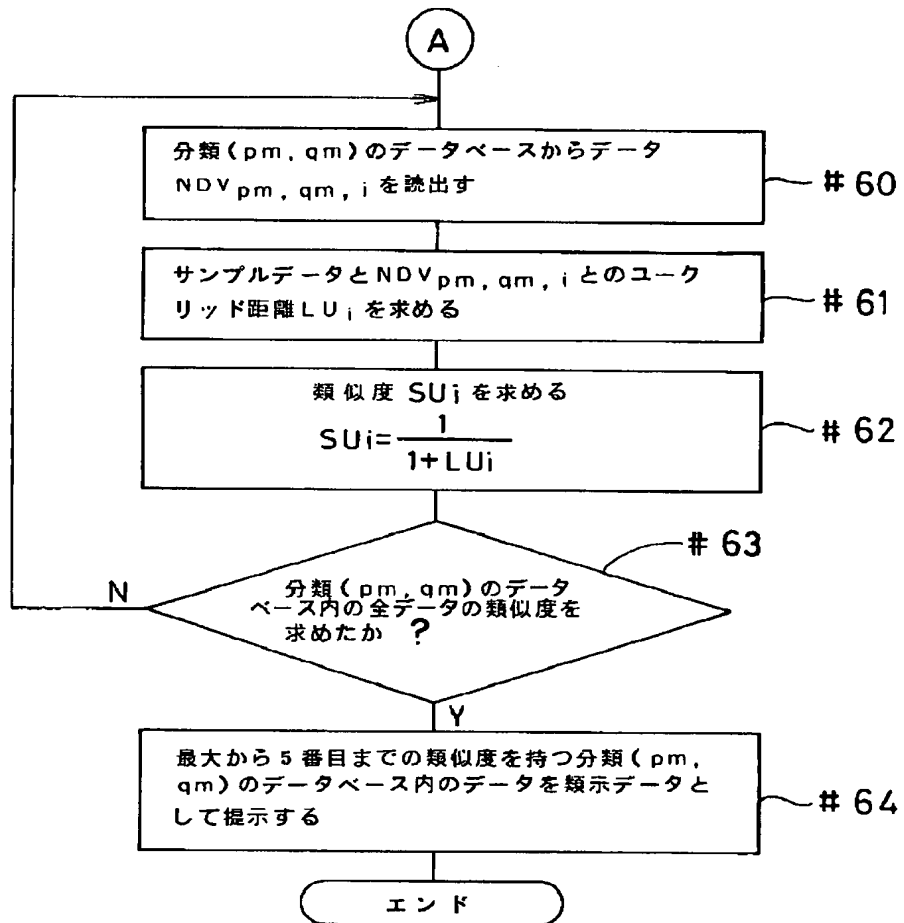
【図8】



【図9】



【図10】



【図 11】

